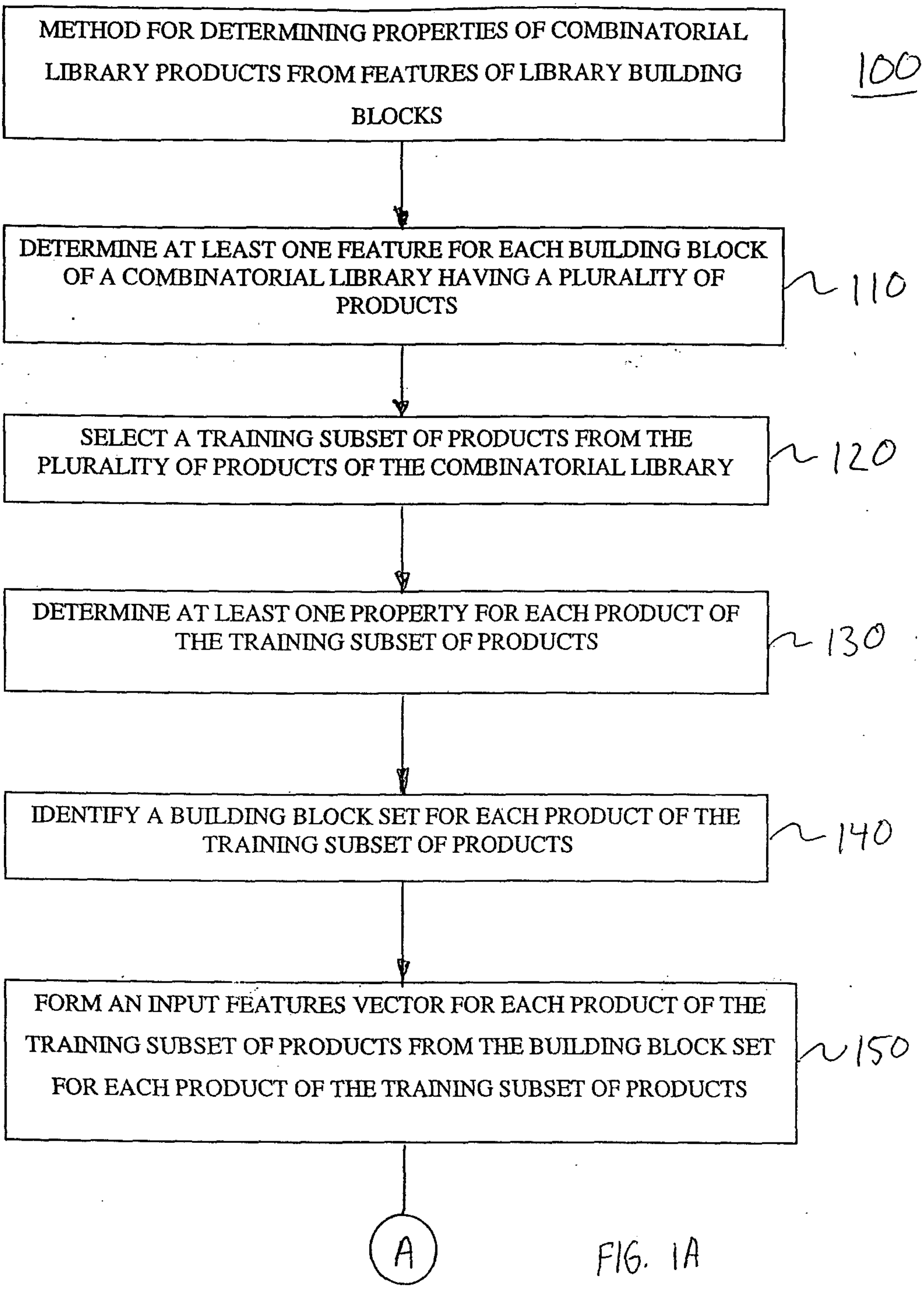
METHOD FOR DETERMINING PROPERTIES OF COMBINATORIAL LIBRARY PRODUCTS FROM FEATURES OF LIBRARY BUILDING BLOCKS
100
DETERMINE AT LEAST ONE FEATURE FOR EACH BUILDING BLOCK OF A COMBINATORIAL LIBRARY HAVING A PLURALITY OF PRODUCTS
110
SELECT A TRAINING SUBSET OF PRODUCTS FROM THE PLURALITY OF PRODUCTS OF THE COMBINATORIAL LIBRARY
120
DETERMINE AT LEAST ONE PROPERTY FOR EACH PRODUCT OF THE TRAINING SUBSET OF PRODUCTS
130
IDENTIFY A BUILDING BLOCK SET FOR EACH PRODUCT OF THE TRAINING SUBSET OF PRODUCTS
140
FORM AN INPUT FEATURES VECTOR FOR EACH PRODUCT OF THE TRAINING SUBSET OF PRODUCTS FROM THE BUILDING BLOCK SET FOR EACH PRODUCT OF THE TRAINING SUBSET OF PRODUCTS
150
A

FIG. 1A

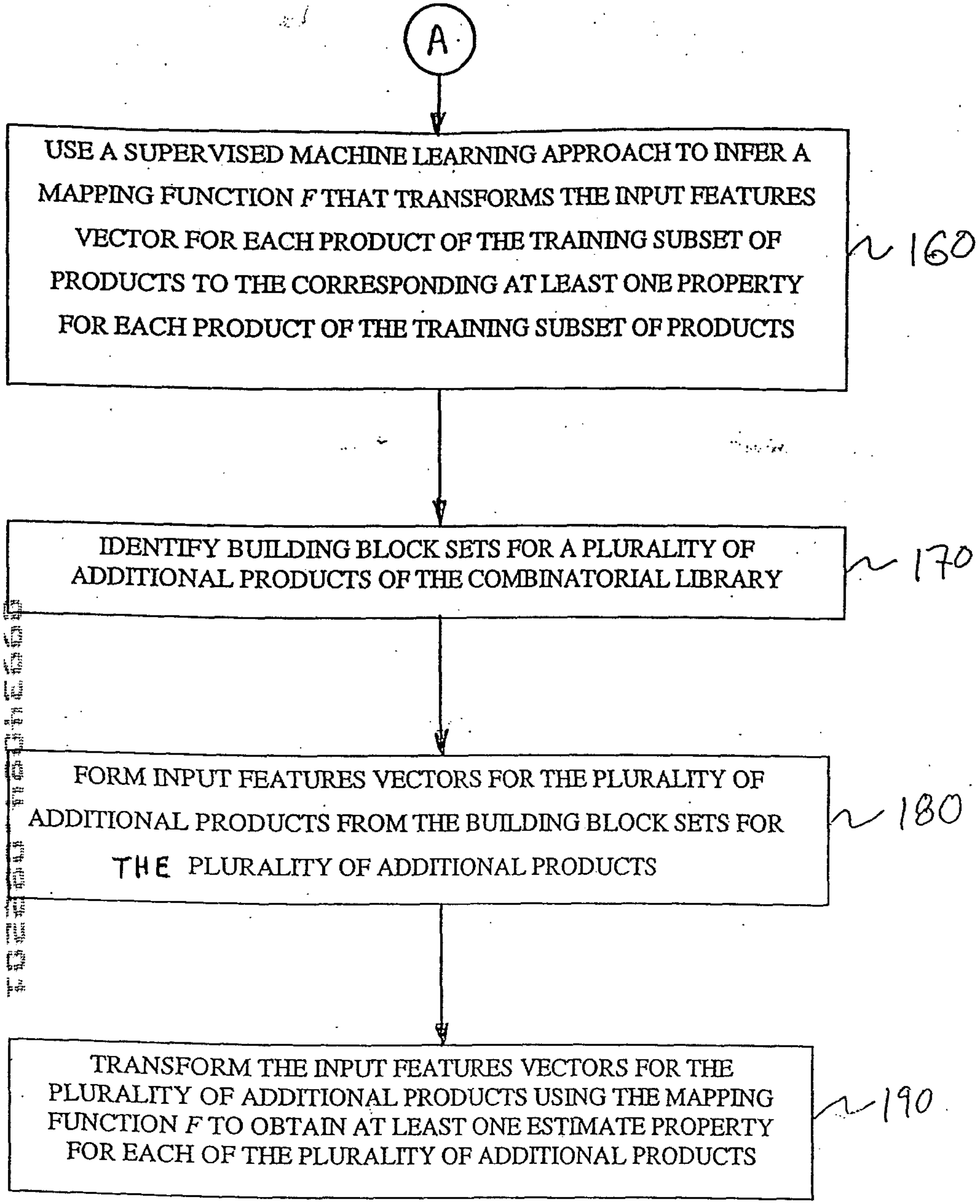
A
USE A SUPERVISED MACHINE LEARNING APPROACH TO INFER A MAPPING FUNCTION *F* THAT TRANSFORMS THE INPUT FEATURES VECTOR FOR EACH PRODUCT OF THE TRAINING SUBSET OF PRODUCTS TO THE CORRESPONDING AT LEAST ONE PROPERTY FOR EACH PRODUCT OF THE TRAINING SUBSET OF PRODUCTS
160
IDENTIFY BUILDING BLOCK SETS FOR A PLURALITY OF ADDITIONAL PRODUCTS OF THE COMBINATORIAL LIBRARY
170
FORM INPUT FEATURES VECTORS FOR THE PLURALITY OF ADDITIONAL PRODUCTS FROM THE BUILDING BLOCK SETS FOR THE PLURALITY OF ADDITIONAL PRODUCTS
180
TRANSFORM THE INPUT FEATURES VECTORS FOR THE PLURALITY OF ADDITIONAL PRODUCTS USING THE MAPPING FUNCTION *F* TO OBTAIN AT LEAST ONE ESTIMATE PROPERTY FOR EACH OF THE PLURALITY OF ADDITIONAL PRODUCTS
190

FIG. 1B

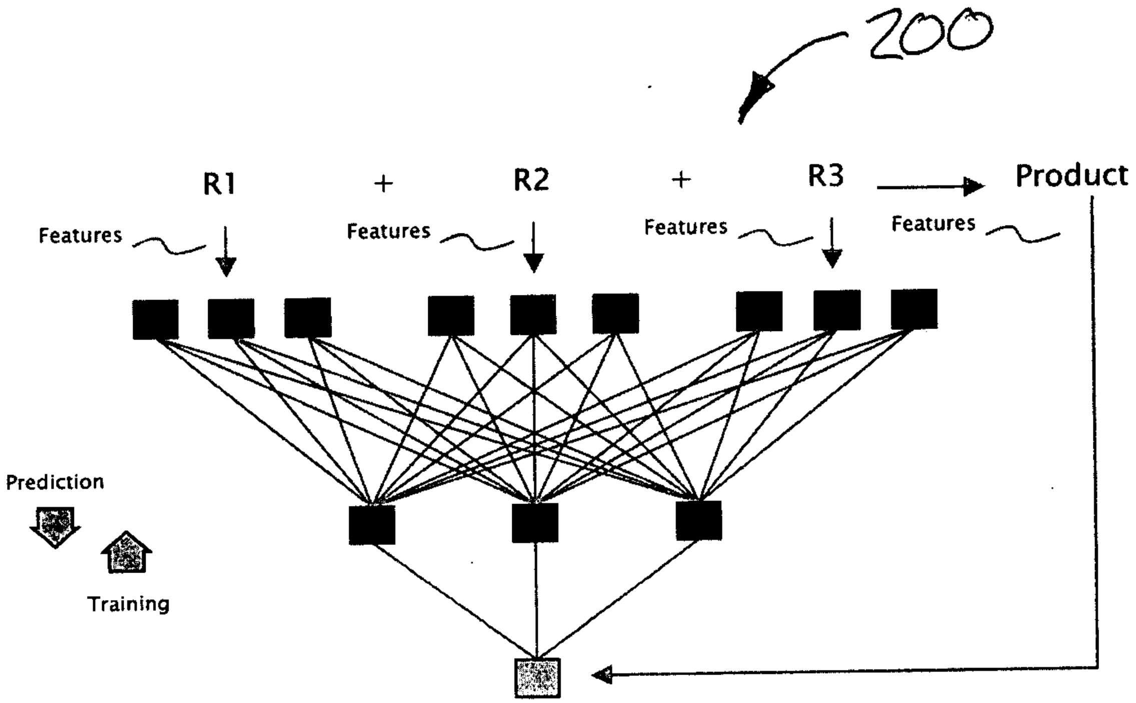
200
R1
+
R2
+
R3
Product
Features
Features
Features
Features
Prediction
Training

FIG. 2

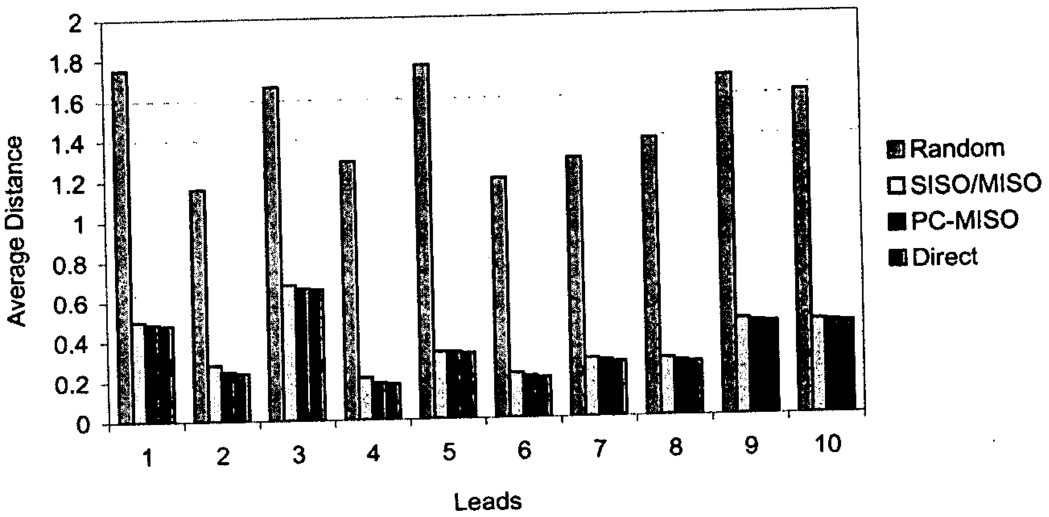
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Average Distance
1
2
3
4
5
6
7
8
9
10
Leads
Random
SISO/MISO
PC-MISO
Direct

FIG. 3

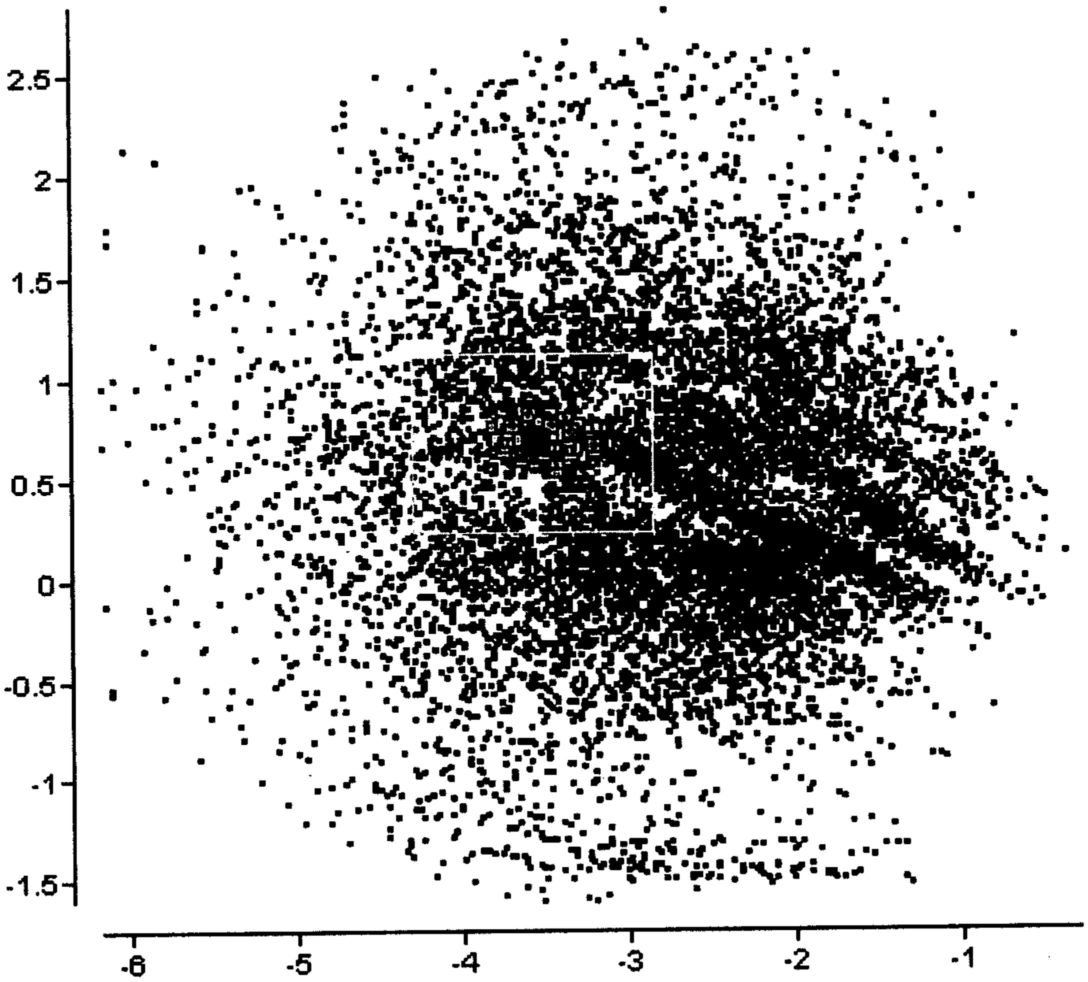

FIG. 4A

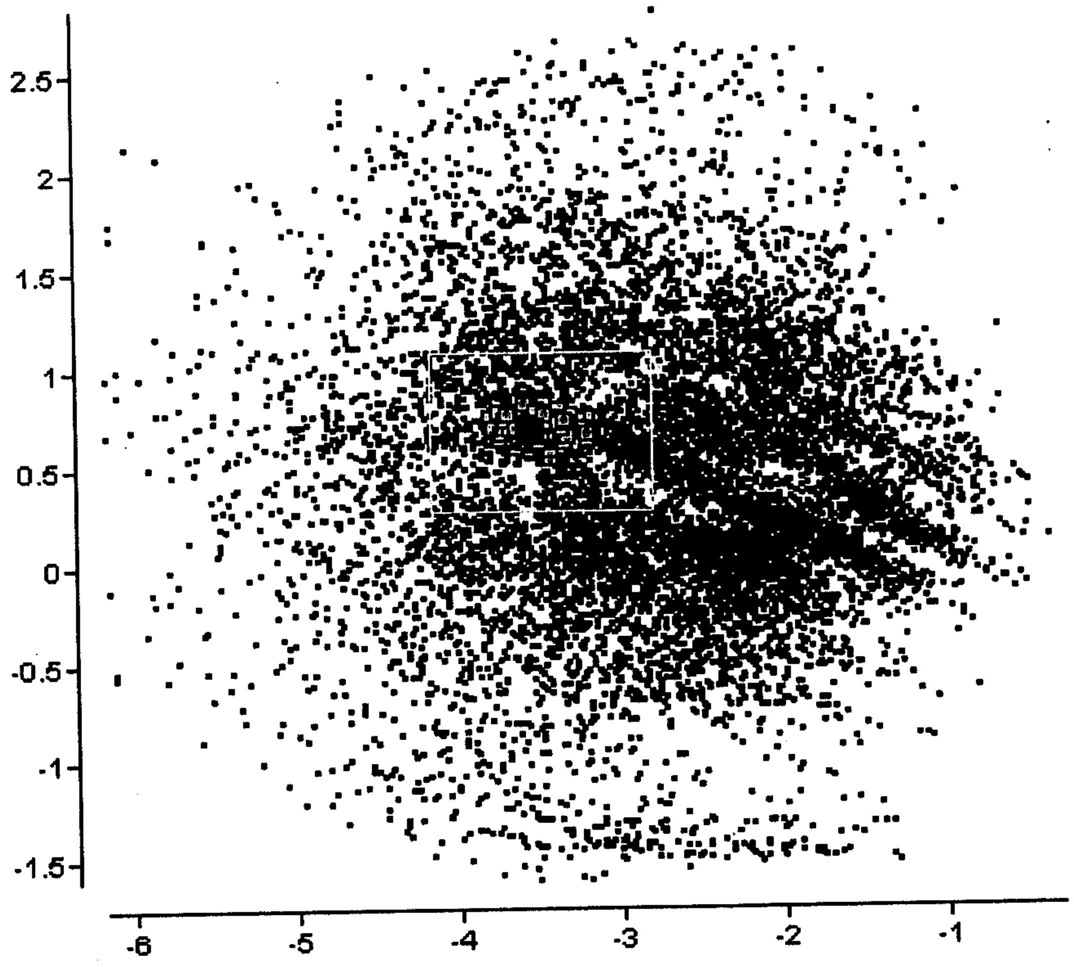

FIG. 4B

09934051.032201

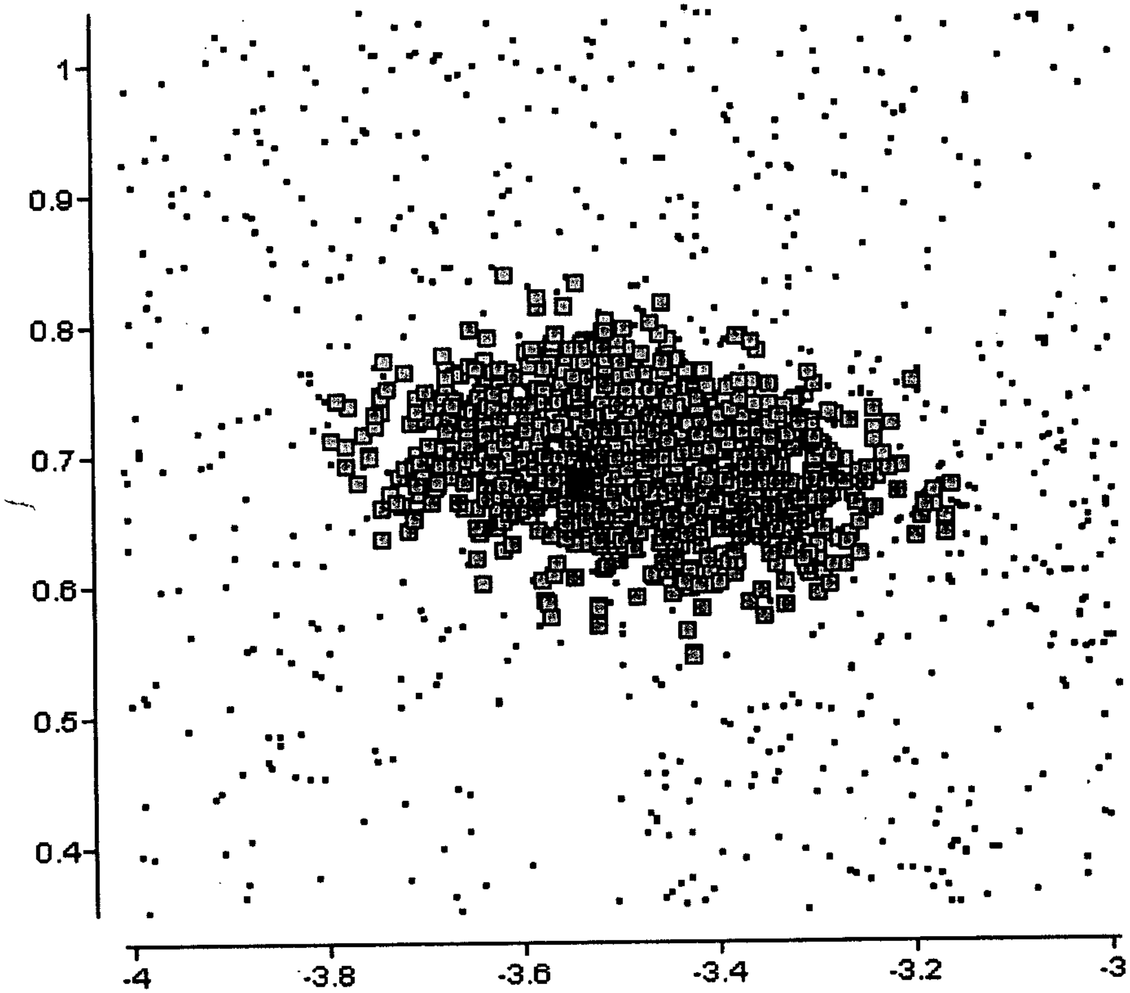

FIG. 4C

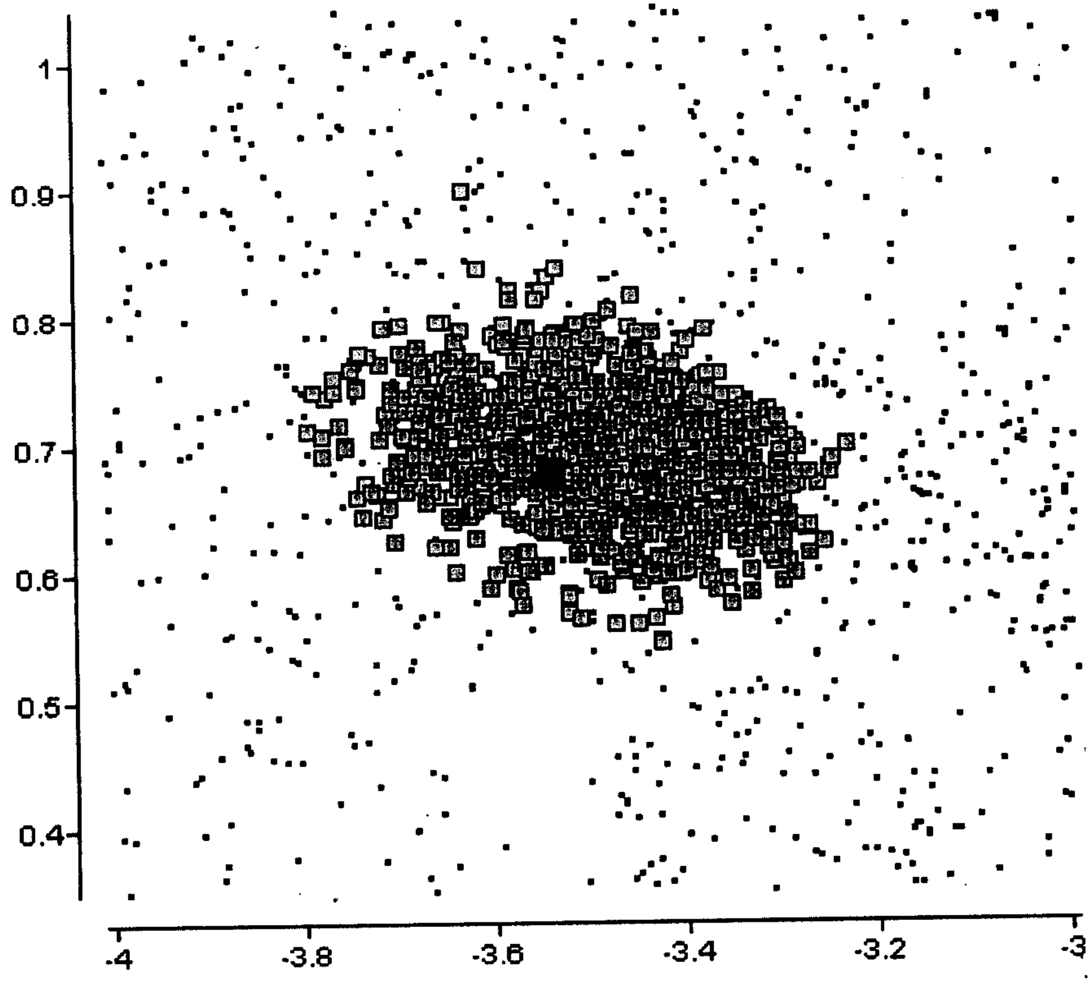

FIG. 4D

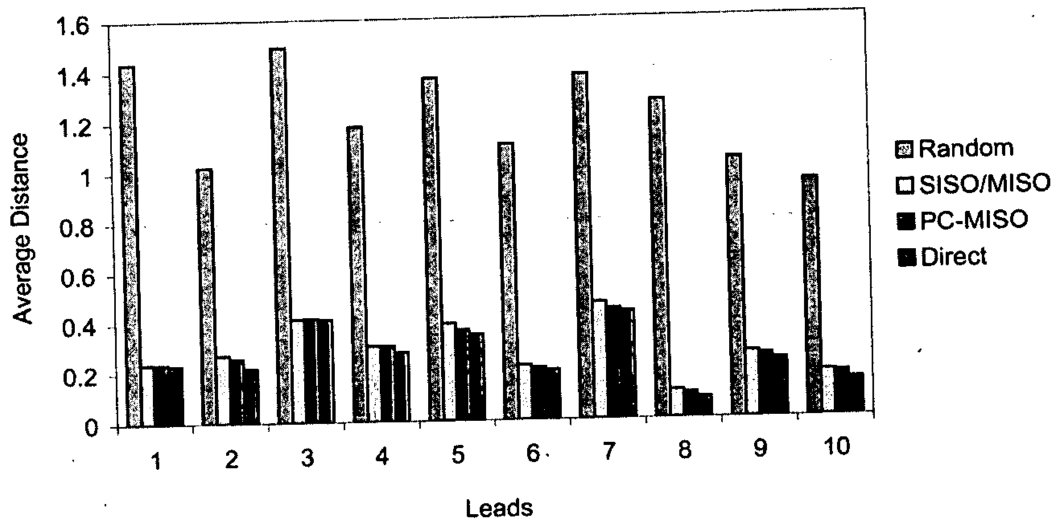
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Average Distance
1
2
3
4
5
6
7
8
9
10
Leads
▨ Random
☐ SISO/MISO
■ PC-MISO
■ Direct

FIG. 5

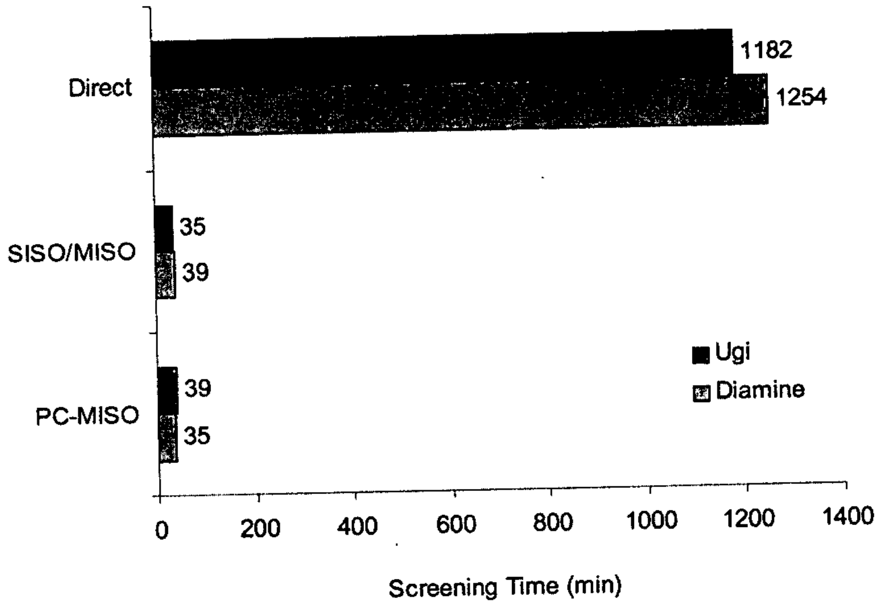
Direct
SISO/MISO
PC-MISO
1182
1254
35
39
39
35
Ugi
Diamine
0
200
400
600
800
1000
1200
1400
Screening Time (min)

FIG. 6

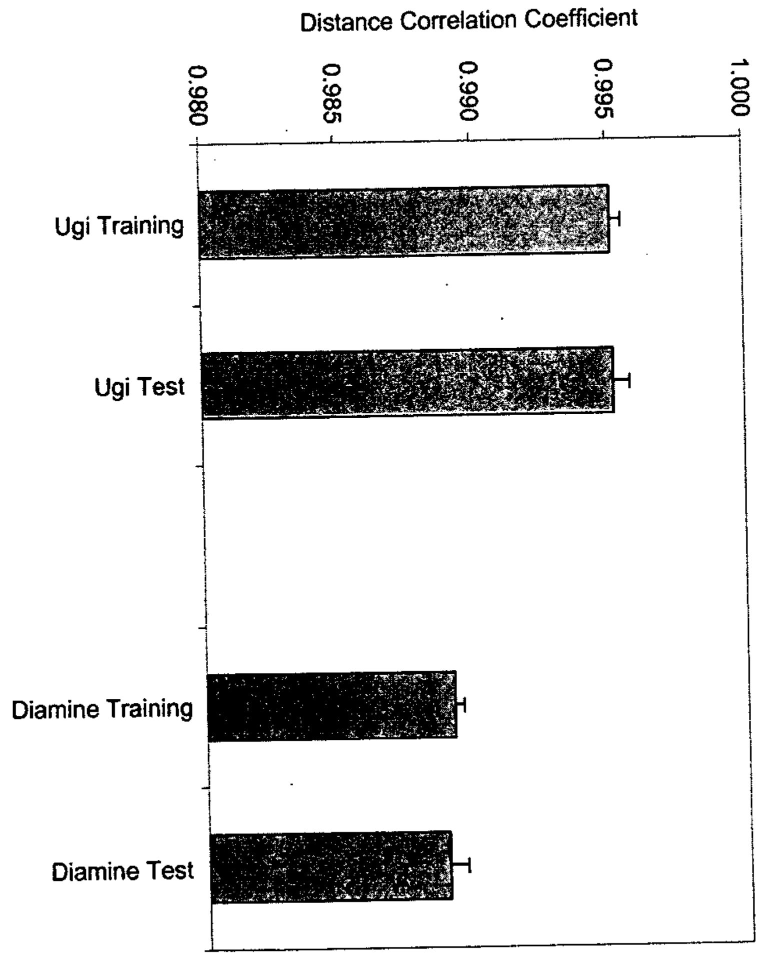

FIG. 7

09934084.062301

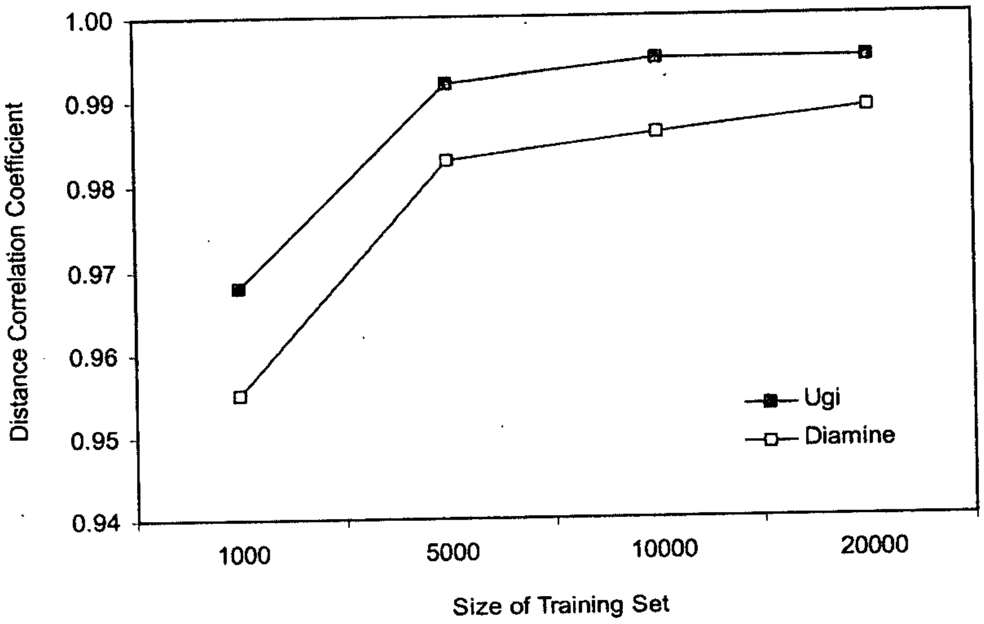
Distance Correlation Coefficient
1.00
0.99
0.98
0.97
0.96
0.95
0.94
1000
5000
10000
20000
Size of Training Set
Ugi
Diamine

FIG. 8

Index	Descriptor	SISO Training R^2	SISO Test R^2	MISO Training R^2	MISO Test R^2
1	No. atoms	0.996	0.997		
2	No. bonds	0.995	0.996		
3	No. elements	0.603	0.614	0.822	0.823
4	Molecular weight	0.996	0.997		
5	Chi 0	0.996	0.997		
6	Chi path 1	0.996	0.997		
7	Chi path 2	0.994	0.995		
8	Chi path 3	0.971	0.973		
9	Chi path 4	0.974	0.976		
10	Chi path 5	0.956	0.957		
11	Chi path 6	0.909	0.910		
12	Chi path 7	0.837	0.843	0.943	0.942
13	Chi path 8	0.666	0.673	0.938	0.934
14	Chi path 9	0.563	0.554	0.939	0.936
15	Chi path 10	0.447	0.457	0.950	0.950
16	Chi cluster 3	0.988	0.987		
17	Chi cluster 4	0.993	0.993		
18	Chi path/cluster 4	0.978	0.980		
19	Val chi 0	0.996	0.997		
20	Val chi path 1	0.997	0.998		
21	Val chi path 2	0.996	0.996		
22	Val chi path 3	0.993	0.994		
23	Val chi path 4	0.981	0.982		
24	Val chi path 5	0.952	0.951		
25	Val chi path 6	0.907	0.905		

FIG. 9A

Index	Descriptor	SISO Training R^2	SISO Test R^2	MISO Training R^2	MISO Test R^2
26	Val chi path 7	0.773	0.775	0.901	0.905
27	Val chi path 8	0.619	0.621	0.890	0.889
28	Val chi path 9	0.349	0.328	0.910	0.910
29	Val chi path 10	0.222	0.201	0.921	0.920
30	Val chi cluster 3	0.994	0.994		
31	Val chi cluster 4	0.993	0.993		
32	Val chi path/cluster 4	0.988	0.989		
33	Chi chain 3	1.000	1.000		
34	Chi chain 4	1.000	1.000		
35	Chi chain 5	0.979	0.978		
36	Chi chain 6	0.995	0.995		
37	Chi chain 7	0.999	0.999		
38	Chi chain 8	1.000	1.000		
39	Chi chain 9	0.999	0.999		
40	Chi chain 10	0.999	0.998		
41	val chi chain 3	1.000	1.000		
42	val chi chain 4	1.000	1.000		
43	val chi chain 5	0.994	0.996		
44	val chi chain 6	0.994	0.995		
45	val chi chain 7	0.998	0.998		
46	val chi chain 8	1.000	1.000		
47	val chi chain 9	0.997	0.998		
48	val chi chain 10	0.986	0.980		
49	subgraph count path 2	0.996	0.997		
50	subgraph count path 3	0.990	0.990		

FIG. 9B

Index	Descriptor	SISO Training R^2	SISO Test R^2	MISO Training R^2	MISO Test R^2
51	subgraph count path 4	0.957	0.960		
52	subgraph count path 5	0.914	0.918		
53	subgraph count path 6	0.837	0.844	0.909	0.905
54	subgraph count path 7	0.752	0.770	0.892	0.887
55	subgraph count path 8	0.582	0.599	0.907	0.906
56	subgraph count path 9	0.446	0.448	0.933	0.932
57	subgraph count path 10	0.366	0.383	0.947	0.945
58	subgraph count cluster 3	0.994	0.995		
59	subgraph count cluster 4	0.991	0.991		
60	subgraph count path/cluster 4	0.980	0.980		
61	subgraph count ring 3	1.000	1.000		
62	subgraph count ring 4	1.000	1.000		
63	subgraph count ring 5	0.995	0.995		
64	subgraph count ring 6	0.994	0.995		
65	subgraph count ring 7	1.000	1.000		
66	subgraph count ring 8	1.000	1.000		
67	subgraph count ring 9	1.000	1.000		
68	subgraph count ring 10	0.999	0.999		
69	kappa 0	0.980	0.980		
70	kappa 1	0.991	0.992		
71	kappa 2	0.907	0.908		
72	kappa 3	0.709	0.710	0.806	0.799
73	kappa alpha 1	0.987	0.987		
74	kappa alpha 2	0.895	0.897	0.960	0.955
75	kappa alpha 3	0.685	0.686	0.774	0.770

FIG. 9C

Index	Descriptor	SISO Training R^2	SISO Test R^2	MISO Training R^2	MISO Test R^2
76	Wiener path no.	0.967	0.965		
77	total Wiener path no.	0.903	0.892		
78	Shannon Index	0.911	0.911		
79	total no. of paths	0.939	0.932		
80	Bonchev-Trinajstić IdW index	0.958	0.955		
81	Bonchev-Trinajstić mean IdW index	0.972	0.972		
82	Bonchev-Trinajstić IdC index	0.979	0.978		
83	Bonchev-Trinajstić mean IdC index	0.793	0.773	0.737	0.759
84	Wiener parity no.	0.988	0.989		
85	Platt F no.	0.996	0.997		
86	Delta partition 1	0.996	0.996		
87	Delta partition 2	0.992	0.992		
88	Delta partition 3	0.997	0.997		
89	Delta partition 4	0.995	0.996		
90	Delta partition 5[1]	1.000	1.000		
91	Delta partition 6[1]	1.000	1.000		
92	No. H	0.996	0.997		
93	No. B[1]	1.000	1.000		
94	No. C	0.997	0.998		
95	No. N	0.995	0.995		
96	No. O	0.994	0.993		
97	No. F	0.996	0.996		
98	No. Si[1]	1.000	1.000		
99	No. P	0.999	0.999		

FIG. 9D

Index	Descriptor	SISO Training R^2	SISO Test R^2	MISO Training R^2	MISO Test R^2
100	*No. S*	0.997	0.999		
101	No. Cl	0.997	0.997		
102	No. Ge^1	1.000	1.000		
103	No. As^1	1.000	1.000		
104	No. Se^1	1.000	1.000		
105	No. Br	1.000	1.000		
106	No. I	1.000	1.000		
107	No. halogens	0.997	0.998		
108	Total topological state 1	0.924	0.918		
109	Total topological state 2	0.947	0.945		
110	Total topological state 3	0.904	0.888		
111	Total topological state 4	0.956	0.956		
112	Total topological state 5	0.852	0.826	0.915	0.907
113	Total topological state 6	0.980	0.980		
114	Total topological state 7	0.832	0.790	0.914	0.898
115	Total topological state 8	0.988	0.988		
116	Total topological state 9	0.913	0.909		
117	Total topological state 10	0.922	0.918		

FIG. 9E

Lead	Random	Direct	SISO/MISO	SISO/MISO	PC-MISO	PC-MISO
	Similarity	Similarity	Similarity	Identity	Similarity	Identity
1	1.754	0.480	0.501	69%	0.486	86%
2	1.158	0.238	0.279	56%	0.244	83%
3	1.664	0.655	0.680	64%	0.660	84%
4	1.291	0.179	0.213	60%	0.186	76%
5	1.763	0.327	0.335	82%	0.334	83%
6	1.196	0.201	0.224	58%	0.209	75%
7	1.294	0.274	0.291	72%	0.283	77%
8	1.385	0.268	0.288	73%	0.275	84%
9	1.694	0.464	0.481	74%	0.470	86%
10	1.613	0.460	0.470	79%	0.464	87%

FIG. 10

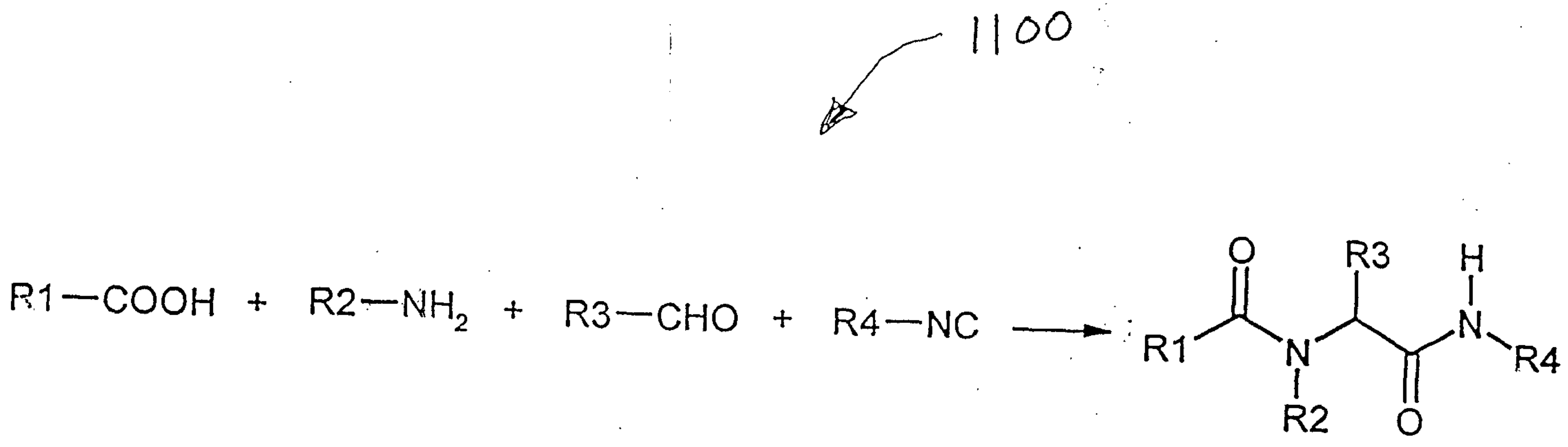

FIG. 11

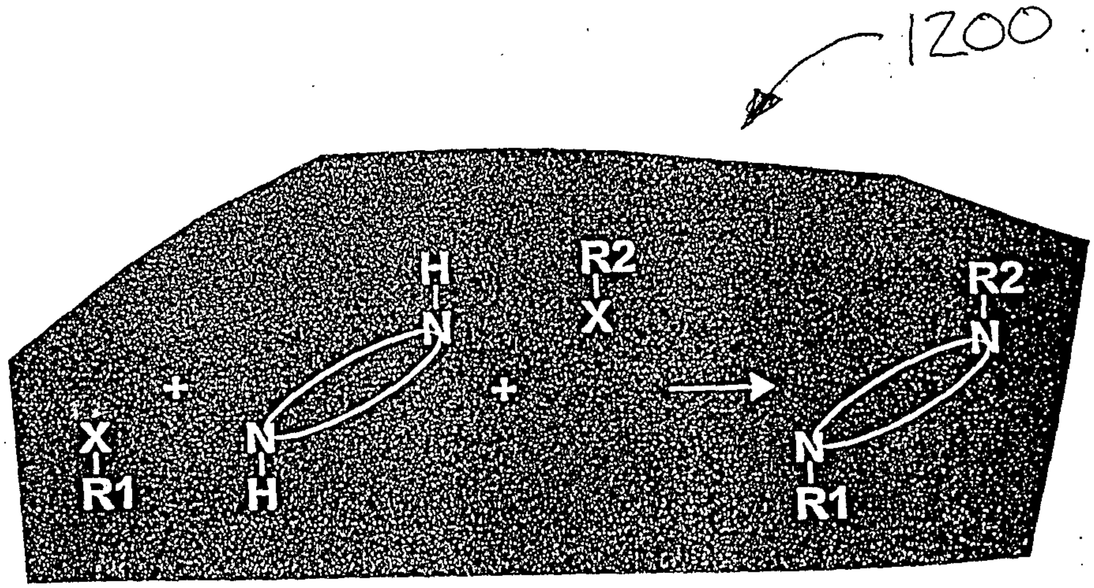

FIG. 12

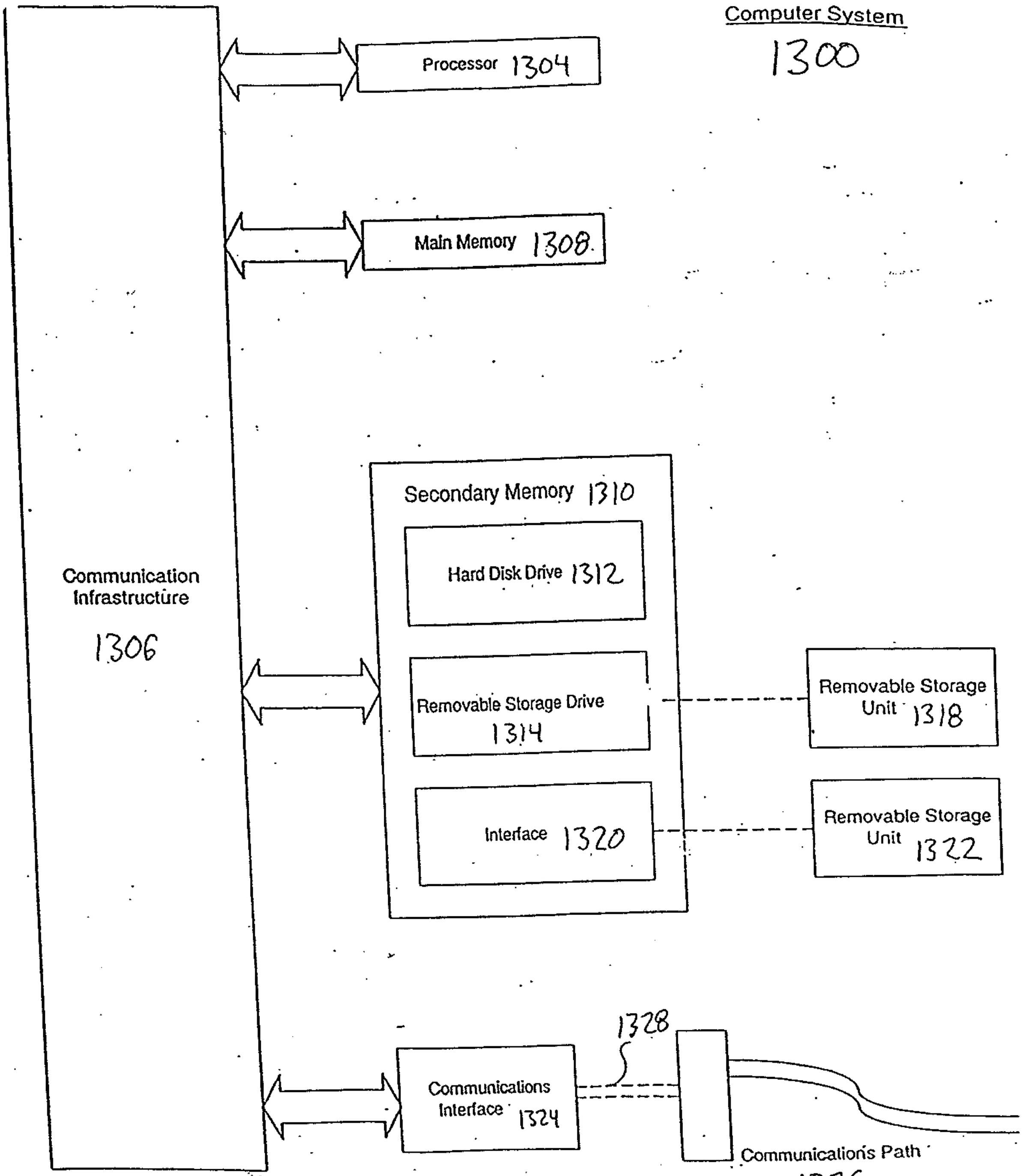
Computer System
1300
Processor 1304
Main Memory 1308
Communication Infrastructure
1306
Secondary Memory 1310
Hard Disk Drive 1312
Removable Storage Drive 1314
Interface 1320
Removable Storage Unit 1318
Removable Storage Unit 1322
1328
Communications Interface 1324
Communications Path
1326

FIG. 13